

Communicating Forensic Evidence: Lessons from Psychological Science

*Barbara A. Spellman**

I. INTRODUCTION	828
II. LEAVE THE STATISTICS AND JARGON TO THE EXPERTS (AND STOP TORTURING THE JURORS)	828
A. Our Numbers Are Not Exact So We Should Not Pretend They Are, Nor Should We Use Non-Statistical Language That Suggests They Are	829
B. The Possibility of Mistake	830
C. Likelihood Ratios and Random Match Probabilities Are Terrible Ways to Present Information to Actual Human Beings	830
D. (Almost) All Evidence is Statistical Anyway	831
III. MISCONCEPTIONS AND EXPLANATION	832
A. Misconceptions Needing Repair	832
B. Explanation in General	833
IV. APPLICATIONS TO UNDERSTANDING LIKELIHOOD RATIOS	834
A. Attribute Substitution	834
B. Explicit Explanation of Variability	836
1. The Importance of Variability	836
2. Form of Explanation for Fingerprints	836
C. Analogy: Explanation Plus Substitution	838
D. Implicit Learning Tasks: Experiencing Variability	839
V. SUMMING UP	840

* University of Virginia School of Law. Thanks to Jay Kochler, Greg Mitchell, and Fred Schauer who gave helpful comments on an earlier draft and to the Australian forensic evidence group who made my visit Down Under intense and informative.

I. INTRODUCTION

In his target article¹ for this symposium celebrating Michael Risinger, Bill Thompson expertly covers a lot of ground regarding the communication of forensic science evidence to laypeople. Among other things, it details and critiques various presentation formats that could be, and have been, used by forensic scientists. It also notes various paradigms that have been used to study how laypeople understand and use that evidence.

In this Article, I focus on alternative forms of explanation that might improve laypeople's understanding of forensic comparison evidence.² As we consider how to best communicate with factfinders, we need to reflect on our beliefs about what they need to understand to make a good decision and what they are likely to understand from various types of presentation. I conclude that rather than attempt to teach factfinders how to do the mathematics involved in forensic statistical reasoning, we should tap into knowledge they already have to develop alternative ways of getting to the "right" conclusions.

The major sections below do the following: Part II describes some reasons why we should not torture jurors (or judges) with lessons in statistics. It refers to some arguments that Thompson made and adds some new ones. Part III notes some misconceptions that laypeople have about forensic analysis and some general principles about the psychology of explanation. Part IV illustrates four possible ways to promote better judgments involving forensic comparison evidence without ever saying "likelihood ratio" or "random match probability" or "Bayes' Theorem": attribute substitution, explanation, analogy, and implicit learning tasks.

II. LEAVE THE STATISTICS AND JARGON TO THE EXPERTS (AND STOP TORTURING THE JURORS)

There are a bunch of reasons why one should not try to explain to a jury (or judge) the excruciating details of likelihood ratios or random match probabilities. And there are also a bunch of reasons not to use the jargon of forensic science.

¹ William C. Thompson, *How Should Forensic Scientists Present Source Conclusions?*, 48 SETON HALL L. REV. 773 (2018).

² I had wanted to allude to the Far Side cartoon (See Gary Larson, *Cartoon of What We Say to Dogs*, THE FAR SIDE COMIC STRIP, <https://www.pinterest.com/pin/265853184223802510/>), depicting a man lecturing his dog about staying out of the garbage while the dog hears merely a lot of noise interspersed with her name (Ginger). However, Dawn McQuiston-Surrett and Michael J. Saks beat me to it with the title of their paper: *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, 33 LAW & HUM. BEHAV. 436 (2009).

A. *Our Numbers Are Not Exact So We Should Not Pretend They Are, Nor Should We Use Non-Statistical Language That Suggests They Are*

Forensic scientists are often concerned with conveying very large (or very small) exact numbers—typically probabilities or likelihoods—to jurors. Many studies using vignettes (or videos) with between-subject mock juror judgments, where different people have different amounts of information, show that jurors do not use such forensic evidence in a statistically consistent or appropriate manner.³ At least for now, we should not worry about such findings. First, as Thompson notes in his target article,⁴ the science behind the numbers is not so well-developed that we can be sure the numbers are even close to accurate. Second, people are likely to overweight seemingly precise numerical values, at least when they are presented by a human communicator (rather than by, for example, a computer message).⁵

Rather than use numbers, forensic scientists often use words like “individualization” and “match.” As Thompson notes, these are dangerous terms, and are likely to be overvalued.⁶ The problem is that laypeople (sensibly) believe that these terms mean there is exactly one person whose prints they could be. In a laboratory study that varied fingerprint examiner terminology, participants who were told that a fingerprint was “individualized,” were much more likely to think the defendant was the one who left the prints at the scene than were participants who were told that the fingerprint was individualized but that “it is possible that the print in question could have come from someone else.”⁷

³ See generally Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859 (1996); Jason Schklar & Shari Seidman Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 LAW & HUM. BEHAV. 159 (1999); Nicholas Scurich & Richard S. John, *Mock Jurors’ Use of Error Rates in DNA Database Trawls*, 37 LAW & HUM. BEHAV. 424 (2013). But see William C. Thompson & Eryn J. Newman, *Lay Understanding of Forensic Statistics: Evaluation of Random Match Probabilities, Likelihood Ratios, and Verbal Equivalents*, 39 LAW & HUM. BEHAV. 332 (2015) (presenting data showing that mock jurors’ probabilistic judgments sometimes approximate Bayesian norms).

⁴ Thompson, *supra* note 1, at 778–81.

⁵ See Y. Charles Zhang & Norbert Schwarz, *The Power of Precise Numbers: A Conversational Logic Analysis*, 49 J. EXPERIMENTAL SOC. PSYCH. 944 (2013).

⁶ Thompson, *supra* note 1, at 797.

⁷ Gregory Mitchell & Brandon Garrett, *The Impact of Proficiency Testing on the Weight Given to Fingerprint Evidence* (Nov. 1, 2017) (unpublished manuscript) (on file with the Seton Hall Law Review). See also Jonathan J. Koehler, *When are People Persuaded by DNA Match Statistics?*, 25 LAW & HUM. BEHAV. 493 (2001) (showing that presenting information using frequencies rather than probabilities makes it easier for people to imagine that the print came from someone other than the suspect). The recognition that someone else could be the source of fingerprints leads some people to fall prey to the “Defense Attorney’s Fallacy”—the belief that if multiple people could be the source, then the evidence is worthless. See William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in*

Using legal terms that have a different meaning from their common non-legal meaning can create problems. For example, mock jurors' prior knowledge of crimes (e.g., believing, falsely, that to be guilty of burglary a person must be armed and must take property) will influence how they label a crime vignette, even when judges provide instructions with the full legal definition. Telling mock jurors to disregard their prior knowledge did not help much; however, instructions that reviewed their false beliefs and tried to replace them with true information did improve decisions.⁸

B. *The Possibility of Mistake*

Worrying about communicating precise small numbers indicating a "match" is unnecessary due to the always-greater probability of an error somewhere in the obtaining, handling, or evaluation of the sample. Although study participants do not always integrate error information well, when participants learned about the chances of a coincidental match and the chances of two types of mistakes—lab error and planted evidence—they did use Bayesian updating in assessing the source probability for DNA evidence (although not for shoeprint evidence).⁹ Participants have also used knowledge of results of an examiner's proficiency testing to adjust the weight given to fingerprint evidence.¹⁰

C. *Likelihood Ratios and Random Match Probabilities Are Terrible Ways to Present Information to Actual Human Beings*

Likelihood ratios, random match probabilities, and some other proposed ways of conveying statistical information to factfinders rely on computing the probability of obtaining the particular evidence given a particular hypothesis. It is futile to try to explain anything that involves $p(e|h)$ —the probability of evidence given a hypothesis—to anyone in an hour, or in a day, or even in a semester of an undergraduate college course in Research Methods and Statistics.

Surveys of psychology academics who rely on p -values for decisions about significance reveal that a large percentage (90% of those not teaching methods) makes common errors when trying to explain what a p -value

Criminal Trials: The Prosecutors' Fallacy and the Defense Attorney's Fallacy, 11 LAW & HUM. BEHAV. 167 (1987).

⁸ See Vicki L. Smith, *When Prior Knowledge and Law Collide: Helping Jurors Use the Law*, 17 LAW & HUM. BEHAV. 507 (1993); Vicki L. Smith, *Prototypes in the Courtroom: Lay Representations of Legal Concepts*, 61 J. PERSONALITY & SOC. PSYCHOL. 857 (1991).

⁹ See Thompson & Newman, *supra* note 3.

¹⁰ See Mitchell & Garrett, *supra* note 7. The experience of the examiner was also shown to matter in Jonathan J. Koehler, N. J. Schweitzer, Michael J. Saks & Dawn McQuiston, *Science, Technology, or Examiner Experience: What Influences Jurors' Judgments about Forensic Science Testimony?*, 22 PSYCHOL. PUB. POL'Y & L. 401 (2016).

represents. They are likely to say that p is the probability that given the data, the null hypothesis (of no difference between means), is true. That is (in a world of two exclusive/exhaustive hypotheses), they believe that $1-p$ represents the likelihood that their own hypothesis is true given the data—a form of $p(h|e)$. But . . . alas, no. A p -value represents the probability of getting the data (or, rather, data that is as extreme or more extreme than the data obtained) given that the null hypothesis is true—a form of $p(e|h)$. If academic researchers who have worked with p -values for years, and are invested in understanding them, cannot get it right, there is little hope for factfinders.¹¹

The problem is, of course, that $p(e|h)$ is not what people want to know. The statistic gets (unintentionally) reinterpreted, typically by way of the Prosecutor's Fallacy¹² or the Source Probability Error¹³ into something that factfinders believe is useful.

D. (Almost) All Evidence is Statistical Anyway

The debate about whether jurors should have to deal with probabilities/statistics is not one worth having. Jurors already do. The reliability of every piece of evidence and every witness is in play—and we can think of that as invoking jurors' statistical sensibilities. Is this witness *likely* to be lying under these circumstances? How likely? What are the chances that installing a safety device would have prevented the accident? We accept that jurors must deal with this kind of uncertainty.¹⁴ We also ask them to make judgments that are “more likely than not”—which certainly sound like we expect them to assess probabilities.

Why do we think forensics is special? Perhaps we think it is special because it is “scientific” and jurors typically do not understand how it works. Perhaps we think it is special because it involves some numerical concepts and we know that many people are innumerate or afraid of math.

But maybe instead of worrying so much about how to present exact statistics, we should think more about what people already know and how to leverage it so that they can make better judgments when dealing with

¹¹ The data are from Heiko Haller & Stefan Krauss, *Misinterpretations of Significance: A Problem Students Share with their Teachers*, 7 METHODS PSYCHOL. RES. 1 (2002). For a broader discussion and overview of the relevant data, see Blakeley B. McShane & David Gal, *Statistical Significance and the Dichotomization of Evidence*, 112 J. AM. STAT. ASS'N 885 (2017). McShane and Gal misleadingly describe the Haller and Krauss study as if it had been done by Gerd Gigerenzer in *Mindless Statistics*, 33 J. SOCIO-ECONOMICS 587 (2004). *Id.*

¹² See Thompson & Schumann, *supra* note 7.

¹³ See Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence*, 34 JURIMETRICS J. 21 (1993).

¹⁴ “[M]ost knowledge, and almost all legal evidence, is probabilistic.” *DePass v. United States*, 721 F.2d 203, 207 (7th Cir. 1983) (Posner, J., dissenting).

forensic statistics.

III. MISCONCEPTIONS AND EXPLANATION

To think about how forensic science might best be communicated to factfinders, we should consider what psychological science has discovered about explaining and teaching complex concepts generally. We should also consider what preconceptions and misconceptions factfinders might have at the start. And, of course, we should recognize that we (e.g., anyone who can read and understand Thompson's target article) are unlikely to be able to gauge what jurors will understand; our intuitions are blown, we suffer from the "curse of knowledge." The "curse of knowledge" refers to a psychological experimental finding that people are likely to misattribute their estimates of the ease of a task (e.g., understanding forensic statistics) to features of the task itself rather than appropriately to their own prior knowledge or experience.¹⁵ To experts, too much seems too easy.

A. *Misconceptions Needing Repair*

People are likely to have several misconceptions about the forensic disciplines. For example, I suspect that people have (at least) four important misconceptions about fingerprints. First, most laypeople believe that every person has unique fingerprints. In two samples, involving more than 1,200 Amazon Turk workers, about 95% of US respondents indicated belief that fingerprints are unique.¹⁶ It might be useful to tell factfinders that forensic scientists do not actually know whether people have unique-as-snowflakes fingerprints.

A second misconception is that if there is a latent print, it can be used to identify a source. People do not seem to realize that there are often partial and degraded prints that do not provide enough information to be useful and that even prints that pass a minimum threshold still vary in the amount of information they contain.

The third and most important misconception is not about between-person variability of prints but rather is about within-person variability of prints. I expect that most people believe that every time a source leaves a fingerprint, it will look remarkably "the same". Although I do not have reliable data, this guess comes not only from the ease with which fingerprint hits occur on television (often with a lovely computer lightshow accentuating

¹⁵ See Colleen M. Kelley & Larry L. Jacoby, *Adult Egocentrism: Subjective Experience Versus Analytic Bases for Judgment*, 35 J. MEMORY & LANGUAGE 157 (1996).

¹⁶ Brandon Garrett & Gregory Mitchell, *How Jurors Evaluate Fingerprint Evidence: The Relative Importance of Match Language, Method Information, and Error Acknowledgment*, 10 J. EMPIRICAL LEGAL STUD. 484 (2013). In another study sample, 84.3% of participants indicated that they believed fingerprints were unique. Mitchell & Garrett, *supra* note 7.

the “match”) but also from years of watching the faces of students when I say to them: “Every time you lay down a fingerprint it is different in some way from every other one you have ever produced.” This fact startles the students and I do give them more explanation (see Part IV, Section B, below).

A fourth misconception—or, perhaps, something most people do not think of unless reminded—is that there are many different potential types of errors in the process—including in procurement, laboratory testing, or reporting of results.

If I am correct, these are specific misconceptions that need to be repaired early during the explanation of fingerprint evidence. But the misconceptions (or true knowledge) that factfinders bring to the courtroom are likely to vary from one forensic discipline to another. For example, Thompson and Newman¹⁷ show that participants responded differently to the same probability information when framed as DNA evidence versus shoeprint evidence. As mentioned above, people cannot help but bring their own pre-existing knowledge to reasoning tasks; it is likely that their experience with shoes as opposed to DNA made them less responsive to the expert’s shoeprint testimony.

B. *Explanation in General*

Plenty of psychology research (so much that I cannot think of a prototypical cite offhand) has shown that people will remember, understand, and be more likely to use information when they are given an explanation for it. It is one thing to tell people *that* something is the case; it is another to give an explanation that shows *why* or *how* it became the case. As a simple example, you could try to teach someone to remember the numbers in this sequence: 1, 8, 27, 64, 125. Memorizing, reciting, or quickly coming up with the sixth number of the sequence is difficult until one is told (or induces) that the sequence is of integers cubed. The explanation (here a rule) makes it possible to regenerate the sequence and connect it to information that is already known.

People are better at solving problems when they understand the rationale (i.e., have an explanation) for the relationship between elements of the problem.¹⁸ Explanations also benefit from having multiple examples

¹⁷ Thompson & Newman, *supra* note 3. Shoeprint testimony was shown to have had little effect on mock jurors’ decisions in Kristy A. Martire, Richard I. Kemp, Ian Watkins, Malindi A. Sayle & Ben R. Newell, *The Expression and Interpretation of Uncertain Forensic Science Evidence: Verbal Equivalence, Evidence Strength, and the Weak Evidence Effect*, 37 LAW & HUM. BEHAV. 197 (2013).

¹⁸ For example, in the now-famous Wason Selection Task, first described in P.C. Wason, *Reasoning About a Rule*, Q.J. EXP. PSYCHOL. 273 (1968), participants are told about a deck of cards, each with a letter on one side and a number on the other. They are shown four cards:

rather than relying on merely one example. From those many examples, people can abstract underlying principles or rules that would apply in other appropriately similar situations.¹⁹

IV. APPLICATIONS TO UNDERSTANDING LIKELIHOOD RATIOS

Below I describe four techniques that could be useful in helping factfinders grasp statistical concepts important to understanding comparison evidence. Note that I do not believe that the exact statistics necessarily need to be reported after this knowledge (or intuition) is developed. Rather, each of these techniques might give factfinders a sense of how to use forensic evidence, presented in whatever form (words or numbers), to appropriately adjust their beliefs.

A. *Attribute Substitution*

Beginning in the 1960's and 1970's, psychological science had to admit that human decision-making was not "rational," in that it does not follow the laws of logic or probability. In the subsequent tidal wave of hundreds of studies, human decision makers were shown to make errors or reveal biases in judgment.²⁰ An important insight, however, was that the errors or biases were not random; rather, they were systematic, thus revealing some principles underlying human reasoning processes.

A good explanation for why and how people give systematic but irrational answers to statistical questions is this: if someone thinks a question is too hard, or does not know the answer, or does not have time or mental resources to answer it, he or she will find an easier related question and answer that. So, for example, when people are asked whether there are more deaths from airplane or automobile accidents per year, and answer the former, it is because they do not know the true answer and so substitute the question: "Which do I hear more about?"²¹ Here, they are substituting

E K 4 7. They are told there is a rule for the cards: if there is a vowel on one side then there is an even number on the other. Then they are asked which cards must be turned over to ensure that the rule is followed. In this standard variant of the task, most participants err by choosing only the E to turn over, and not acknowledging that the 4 must be turned over as well. An analysis of the many variants of the tasks used suggests that understanding the rationale for the rule is key to getting the problem correct. Patricia W. Cheng & J. Holyoak, *Pragmatic Reasoning Schemas*, 17 COGNITIVE PSYCHOL. 391 (1985).

¹⁹ With regard to using multiple examples in analogical reasoning, see Mary L. Gick & Keith J. Holyoak, *Schema Induction and Analogical Transfer*, 15 COGNITIVE PSYCHOL. 1 (1980).

²⁰ For a brief history and overview of this research, see Barbara A. Spellman & Simone Schnall, *Embodied Rationality*, 35 QUEEN'S L.J. 117 (2009). For a compendium of the early research, see *Judgment Under Uncertainty: Heuristics and Biases* (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982).

²¹ The true answer, of course, is that there are more automobile-related deaths per year.

familiarity for information (using the “availability heuristic”).

But, attribute substitution works to give the correct answer much of the time. For example, when asked “which has more earthquakes, California or New York?” substituting “which do I hear more about?” gives the correct answer. Accordingly, maybe there is something that can be used to replace forensic statistics that would give the right answer most of the time.

Although he does not use these terms, or use it for this purpose, Michael Risinger presents an interesting idea about attribute substitution in his article called “Leveraging Surprise.”²² His article suggests that surprise could be used as a way of measuring how much people believe alleged facts. For example, the more you are surprised to learn that a “fact” is false, the more you must have previously believed it. Risinger wants to apply the surprise test to standards of proof. For example, “I’d be extremely EXTREMELY surprised to learn that the defendant was not the perpetrator” might be a good expression of belief beyond a reasonable doubt. Ginther and Cheng²³ report a quick first experiment in which participants read a vignette about a crime and then estimate the probability of guilt (or innocence) and how surprised they would be to find out that the suspect was innocent (or guilty). The data show a high correlation between people’s estimations of probability and surprise.

The relevance to forensic evidence is this: if people reveal a good correlation when expressing surprise and probability, then perhaps surprise would be a good way for forensic scientists to communicate probability. That is, rather than giving a statistic, or even an expression of probability,²⁴ an expression of surprise by a forensic scientist (“I would be extremely EXTREMELY surprised to learn that it was not the suspect’s fingerprint”) might be a good way to present the equivalent of statistical evidence. And, if surprise does not work, perhaps there is some other attribute that would make a good substitution for probabilities.

Aric Jenkins, *Which is Safer: Airplanes or Cars?*, FORTUNE (July 20, 2017), <http://fortune.com/2017/07/20/are-airplanes-safer-than-cars/> (displaying recent statistics about auto-related fatalities for Americans).

²² D. Michael Risinger, *Leveraging Surprise: What Standards of Proof Imply That We Want from Jurors, and What We Should Say to Them to Get It*, 48 SETON HALL L. REV. (forthcoming 2018).

²³ Matthew Ginther & Edward K. Cheng, *Surprise vs. Probability as a Metric for Proof*, 48 SETON HALL L. REV. (forthcoming 2018).

²⁴ Much research shows that people vary greatly in their interpretation of words like: probably, likely, etc. See David V. Budescu & Thomas S. Wallsten, *Processing Linguistic Probabilities: General Principles and Empirical Evidence*, 32 PSYCHOL. LEARNING & MOTIVATION 275 (1995).

B. *Explicit Explanation of Variability*

Attribute substitution relies on people's unconscious strategy to answer questions using a method they already know. Good explanations that draw on people's pre-existing knowledge of category variability might also help them understand forensic conclusions.

1. The Importance of Variability

It turns out that factfinders are likely to already know about the importance of variability within a category for making inferences about that category. In a classic study, participants were told to imagine that they were explorers who had discovered a new island in the Southeastern Pacific Ocean. They encounter a member of the native population who has brown skin and is obese, a new type of bird that is blue in color and builds its nest in a eucalyptus tree, and a new element that conducts electricity and burns with a green flame. When asked what percentage of the members of each category (human/bird/element) they thought would share the sample property, the answers ranged from very few (obesity) to nearly all (conduct electricity, burn with green flame). Why? Because the participants understood the underlying variability in those properties.²⁵ Basic knowledge about category variability is useful but it depends on kinds of background knowledge of the category that people are not likely to have about fingerprints.

2. Form of Explanation for Fingerprints

When I talk about variability in fingerprints to my class, I say something like the following. It is meant to evoke experiences and pre-existing knowledge, describe the underlying "how" of fingerprint creation and use, and impart (I hope) an understanding of why there is not a one-and-only perfect match.

How many of you have ever had your fingerprints taken?²⁶ If it was a while back, like when I first had mine done, there was a guy grasping my finger making sure that I rolled it evenly in the ink, and then that I used a constant pressure as I rolled my finger on the print paper.²⁷ These days, of course, you place your fingers

²⁵ Richard E. Nisbett, David H. Krantz, Christopher Jepson & Ziva Kunda, *The Use of Statistical Heuristics in Everyday Inductive Reasoning*, 90 PSYCHOL. REV. 339 (1983). The research is also described in JOHN H. HOLLAND, KEITH J. HOLYOAK, RICHARD E. NISBETT & PAUL R. THAGARD, *INDUCTION: PROCESSES OF INFERENCE, LEARNING, AND DISCOVERY* (1986).

²⁶ Currently, among the law students at the University of Virginia, it is nearly everyone. In the olden days, I used to joke that they did not have to admit to having had their fingerprints taken and some students seemed relieved.

²⁷ Acting this out with body contortions is appreciated by the audience.

spread out, but not too much, on a scanner, and it tells you when you have good placement for the scan. Okay, fine. But now let's consider your fingerprint as picked up by the cops—from a glass or a doorknob or a gun—which has been smudged by other later prints, or is on an uneven surface, or you touched or grabbed briefly while twisting as you ran. That print, which ends up in the lab after the examiner “picks it up” (in some way that may also change the print), does not, cannot, look exactly the same as your pristine carefully placed original print. In fact, even if you were to go back and *try your best to create the exact same original print*, you could get very close but it would not be exact. So, your fingerprints vary from one time to the next—sometimes a little and sometimes a lot.

Another thing about fingerprints is that although you may often hear that “everyone’s fingerprints are different,” we do not know that for a fact. We certainly do not have the fingerprints of everyone in the world on file to compare. Plus, sometimes we only have partial prints. And we certainly cannot rule out that some *parts* of someone’s fingerprint might be identical to some *parts* of someone else’s fingerprint.

But even if we could say that everyone’s prints are different, we might still be in trouble because we do not know whether they are *different enough* to distinguish them. Because if your fingerprints differ every time and place you leave them, and other people’s fingerprints differ every time they leave them, how do we know that even if your actual fingers are different, that the prints you end up leaving might not look close to identical? And that’s the problem—there is a fingerprint there on the gun—and it looks like it could be from my finger or from yours.²⁸

Perhaps something like this could trigger better reasoning about the comparison disciplines (e.g., both the ones in which people are less experienced like fingerprints or ballistics, and more experienced like shoeprints or handwriting). Then each discipline could supplement training with domain-specific information.

²⁸ For a description of between versus within variability, see WILLIAM THOMPSON, JOHN BLACK, ANIL JAIN & JOSEPH KADANE, AAAS, FORENSIC SCIENCE ASSESSMENTS: A QUALITY AND GAP ANALYSIS—LATENT FINGERPRINT EXAMINATION 17–27 (Sept. 15, 2017), https://mcmprodaas.s3.amazonaws.com/s3fs-public/reports/Latent%20Fingerprint%20Report%20FINAL%209_14.pdf?i9xGS_EyMHnIPLG6INIUYzb66L5cLdlb.

1. *Analogy: Explanation Plus Substitution*

Another way to tap into using peoples' pre-existing knowledge of variability to bolster their understanding of likelihood ratios in forensic evidence is to use analogy. Using analogy has characteristics of both explanation and attribute substitution. To use analogy to communicate to another person, we need to find a domain that is analogous to forensic science statistics (the "target" domain) in ways we care about—e.g., has likelihoods, base rates, variability, and uncertainty—but that the learner has prior experience with, and, preferably, might even have some pre-existing expertise (the "source" domain).²⁹ Something like face recognition³⁰ might work.

Step one: Describe some basic important features of the target domain—forensic comparison processes—that might be unknown to the listener. (For fingerprints, see example description in Part IV, Section B, above.)

Step two: Remind the learner of important (relevant) aspects of the source domain. Something like: "Now comparing fingerprints seems very complicated, but it is quite similar to something you do every day, something you are an expert at, but do not even think about much—and that is recognizing faces. You know what *you* look like, but you are probably aware that when you look in the mirror from one day (or one hour) to the next, you do not look exactly the same. There are bad hair days, bags-under-eyes days, what's-wrong-with-my-skin days, and all sorts of intermediate variations. Yet, you do not have any trouble recognizing yourself in the mirror. Nor do you have trouble recognizing your close friends, or family, or people you see daily even though, in fact, they look different every day."

Step three: Draw out the important similarities for understanding the target domain. "But now consider someone you do not know well or have not seen in a while who you run into at a party or, worse yet, a college reunion. You remember what Nikki looked like 30 years ago, but is that person Nikki or Jane? Would someone who looked like Nikki 30 years ago look like this person in front of you now? Or would she look entirely different? This is one side of the problem of variability—she could look

²⁹ Mary L. Gick & Keith J. Holyoak, *Analogical Problem Solving*, 12 COGNITIVE PSYCHOL. 306 (1980); Laura R. Novick, *Analogical Transfer, Problem Similarity, and Expertise*, 14 J. EXPERIMENTAL PSYCHOL.: LEARNING, MEMORY & COGNITION 510 (1988). For a description of these and related analogy studies as relevant to judicial decision-making, see Barbara A. Spellman, *Judges, Expertise, and Analogy*, in THE PSYCHOLOGY OF JUDICIAL DECISION-MAKING 149 (David E. Klein & Gregory Mitchell eds., 2010) and Barbara A. Spellman and Frederick Schauer, *Analogy, Expertise, and Experience*, 84 U. CHI. L. REV. 249 (2017).

³⁰ Of course, face recognition could be a tricky example to use because it might be relevant to the case at hand. Perhaps snowflakes?

many different ways now—some easy to recognize, some not. Then there is the other side of the problem—could it be someone *else* who looks like this person in front of you now? Probably yes, if Nikki did not have a lot of distinctive characteristics; but probably no if she had one blue eye, one brown eye, and was six foot ten.”

This exposition provides the building blocks for understanding likelihood ratios without ever saying “likelihood ratio.” And it is easy to add information about how the number of people at the reunion (i.e., the size of the relevant comparison population) would affect one’s belief about how likely it could be someone other than Nikki.

2. *Implicit Learning Tasks: Experiencing Variability*

This last suggestion for helping people understand forensic statistics is the most speculative. It involves a learn-by-doing task that should not need verbal explanation of forensics at the outset.

In the 1980’s, psychologists ran many different experimental tasks to try to understand how people go from seeing instances of categories (e.g., lots of birds, lots of hypothetical Martians) to inducing mental representations of the average member and the variability within the categories.³¹ One set of studies asked participants to imagine two artists: Smith and Wilson. They each designed abstract art that consisted of black and white squares (on a 10x10 grid). The participants’ task was to distinguish the work of the two artists. The “artwork” was created by: (a) creating one basic non-viewed 10x10 grid of black/white squares for each artist; these represented each artist’s prototypical work; then (b) distorting each prototype by changing some percentage of squares from black to white and vice versa; these represent the work the artist produces for viewing.³² In some variants of the studies, one artist’s work was all fairly similar (i.e., few squares were changed) whereas the other artist’s work varied a lot (i.e., many squares were changed). After seeing a few examples of the art, choosing which artist they thought it belonged to, and getting feedback on their decision accuracy, participants then saw many more examples and had to guess which artist they thought produced it, this time without feedback.³³

How is this study relevant here? When participants had to categorize the paintings, they acted as if they were using likelihood ratios based on the number of color switches from the prototype. That is, many of the paintings

³¹ For a review, see Douglas L. Medin, *Concepts and Categories*, 44 AM. PSYCHOLOGIST 1469 (1989).

³² Lisbeth S. Fried & Keith J. Holyoak, *Induction of Category Distributions: A Framework for Classification Learning*, 10 J. EXPERIMENTAL PSYCHOL.: LEARNING, MEMORY & COGNITION 234, 241–43 (1984).

³³ *Id.* at 247–48.

could have been made by either artist, but the participants categorized each one based on the likelihood of it being from a particular artist given their knowledge of the variability in the two artists' paintings. Recall that the participants had never seen the artists' prototypes; rather, the variability is induced from the training examples.³⁴

One potential strength of this procedure is that it is more like learning about relative frequencies than about probabilities. In many judgment tasks, people are better at evaluating information when it is presented as frequencies rather than probabilities.³⁵

This study demonstrates implicit learning of category distributions. In most laboratory tasks like this one (i.e., category induction from abstract highly variable examples), participants cannot explain the criteria they are using in making their decisions. I do not know whether explaining to them that what they have learned to do is very similar statistically to what forensic analysts do would help them to understand the statistics or the process. But it might.

V. SUMMING UP

In this Article, I suggest two major things. First, I jump on the bandwagon of doubts about some of the ways forensic comparison evidence is presented to factfinders and I add some psychological backing to some of those doubts. Second, I propose that there are several ways, coming from a variety of areas within cognitive psychology, to teach factfinders about variability, to relate it to things they already know, and to use that as scaffolding to increase understanding of, and competence in using, forensic statistical evidence.

³⁴ *Id.*

³⁵ The usefulness of frequencies is shown often in the work of Gerd Gigerenzer. For application to DNA match statistics, see Koehler, *supra* note 7.